

Total variation minimization for stable multidimensional signal recovery

Deanna Needell* and Rachel Ward†

October 12, 2012

Abstract

Consider the problem of reconstructing a multidimensional signal from partial information. Without any additional assumptions, this problem is ill-posed. However, for signals such as natural images or movies, the minimal *total variation* estimate consistent with the measurements often produces a good approximation to the underlying signal, even if the number of measurements is far smaller than the ambient dimensionality. While reconstruction guarantees and optimal measurement designs have been established for related ℓ_1 -minimization problems, the theory for total variation minimization has remained elusive until recently, when guarantees for two-dimensional images $\mathbf{x} \in \mathbb{C}^{N^2}$ were established. This paper extends the recent theoretical results to signals $\mathbf{x} \in \mathbb{C}^{N^d}$ of arbitrary dimension $d \geq 2$. To be precise, we show that a multidimensional signal $\mathbf{x} \in \mathbb{C}^{N^d}$ can be reconstructed from $\mathcal{O}(sd \log(N^d))$ linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x}$ using total variation minimization to within a factor of the best s -term approximation of its gradient. The reconstruction guarantees we provide are necessarily optimal up to polynomial factors in the spatial dimension d and a logarithmic factor in the signal dimension N^d . The proof relies on bounds in approximation theory concerning the compressibility of wavelet expansions of bounded-variation functions.

1 Introduction

Compressed sensing (CS) is a new signal processing methodology where signals are acquired in compressed form as undersampled linear measurements. The applications of CS are abundant, ranging from radar and error correction to many areas of image processing [13]. The underlying assumption that makes such acquisition and reconstruction possible is that most natural signals are *sparse* or *compressible*. We say that a signal $\mathbf{x} \in \mathbb{C}^p$ is s -sparse when

$$\|\mathbf{x}\|_0 \stackrel{\text{def}}{=} |\text{supp}(\mathbf{x})| \leq s \ll p. \quad (1)$$

Compressible signals are those which are well-approximated by sparse signals. More generally, a signal $\mathbf{x} \in \mathbb{C}^p$ is said to be s -sparse with respect to a basis \mathbf{B} when \mathbf{x} can be represented as a linear combination of s atoms from \mathbf{B} . In the CS framework, we acquire $m \ll p$ nonadaptive linear measurements of the form

*Claremont McKenna College, 850 Columbia Ave, Claremont CA, 91711, email: dneedell@cmc.edu.

†University of Texas at Austin, 2515 Speedway, Austin, Texas, 77842, email: rward@math.utexas.edu. R. Ward has been supported in part by a Donald D. Harrington Faculty Fellowship, Alfred P. Sloan Research Fellowship, and DOD-Navy grant N00014-12-1-0743.

$$\mathbf{y} = \mathcal{M}(\mathbf{x}) + \boldsymbol{\xi},$$

where $\mathcal{M} : \mathbb{C}^p \rightarrow \mathbb{C}^m$ is an appropriate linear operator and $\boldsymbol{\xi}$ is vector modeling additive noise. The theory of CS [6, 16] ensures that under suitable assumptions on the measurement operator \mathcal{M} , a sufficiently compressible signal can be accurately approximated by the signal of minimal ℓ_1 -norm consistent with the measurements,

$$\hat{\mathbf{x}} = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{w}\|_1 \quad \text{such that} \quad \|\mathcal{M}(\mathbf{w}) - \mathbf{y}\|_2 \leq \varepsilon, \quad (L_1)$$

where $\|\mathbf{w}\|_1 = \sum_i |w_i|$ and $\|\mathbf{w}\|_2 = (\sum_i |w_i|^2)^{1/2}$ denote the standard ℓ_1 and Euclidean norms, and ε bounds the noise level $\|\boldsymbol{\xi}\|_2 \leq \varepsilon$. The program (L_1) may be cast as a second order cone program (SOCP) and can be solved efficiently using standard convex programming methods [10, 14].

To guarantee robust recovery of compressible signals via (L_1) , Candès and Tao introduced in [6] the restricted isometry property (RIP) for a measurement operator \mathcal{M} .

Definition 1. A linear operator $\mathcal{M} : \mathbb{C}^p \rightarrow \mathbb{C}^m$ is said to have the restricted isometry property (RIP) of order $s \in \mathbb{N}$ and level $\delta \in (0, 1)$ if

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|\mathcal{M}(\mathbf{x})\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2 \quad \text{for all } s\text{-sparse } \mathbf{x} \in \mathbb{C}^p. \quad (2)$$

Many distributions of random matrices of dimension $m \times p$ are known to generate RIP matrices of order s and level $\delta \leq c < 1$ for $m \approx \delta^{-2}s(\log p)^4$. Representative families of random matrices include randomly subsampled rows from the discrete Fourier transform [34] or from a bounded orthonormal system more generally [34, 33, 31, 32, 2], and randomly-generated circulant matrices [21]. Moreover, a matrix whose entries are independent and identical (i.i.d.) realizations of a properly-normalized subgaussian random variable will have the RIP with probability exceeding $1 - e^{-cm}$ once $m \approx \delta^{-2}s \log(p/s)$ [7, 27, 34, 1].

Candès, Romberg, and Tao [5] showed that when the measurement operator \mathcal{M} has the RIP of order $\mathcal{O}(s)$ and sufficiently small constant δ , the program (L_1) recovers an estimation $\hat{\mathbf{x}}$ to \mathbf{x} that satisfies the error bound

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C \left(\frac{\|\mathbf{x} - \mathbf{x}_s\|_1}{\sqrt{s}} + \varepsilon \right), \quad (3)$$

where \mathbf{x}_s denotes the best s -sparse approximation to the signal \mathbf{x} . Using properties about Gel'fand widths of the ℓ_1 ball due to Kashin [19] and Garnaev–Gluskin [17], this is the optimal minimax reconstruction rate from $m \approx s \log(p/s)$ nonadaptive linear measurements. Due to the rotational-invariance of an RIP matrix with randomized column signs [22], a completely analogous theory holds for signals that are compressible with respect to a known orthonormal basis or tight frame \mathbf{D} by replacing \mathbf{w} with $\mathbf{D}^*\mathbf{w}$ inside the ℓ_1 -norm of the minimization problem (L_1) [3].

1.1 Imaging with CS

Natural images are highly compressible with respect to their gradient representation. A typical grayscale digital image, regarded as a signal $\mathbf{x} \in \mathbb{C}^{N^2}$, consists primarily of slowly-varying pixel intensities, with large jumps in intensity occurring only along edges. Figure 1 illustrates the gradient sparsity of a representative image $\mathbf{x} \in \mathbb{C}^{N^2}$ along with its discrete directional derivatives,

$$\mathbf{x}_u : \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{(N-1) \times N}, \quad (\mathbf{x}_u)_{j,k} = \mathbf{x}_{j+1,k} - \mathbf{x}_{j,k} \quad (4)$$

$$\mathbf{x}_v : \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times (N-1)}, \quad (\mathbf{x}_v)_{j,k} = \mathbf{x}_{j,k+1} - \mathbf{x}_{j,k} \quad (5)$$

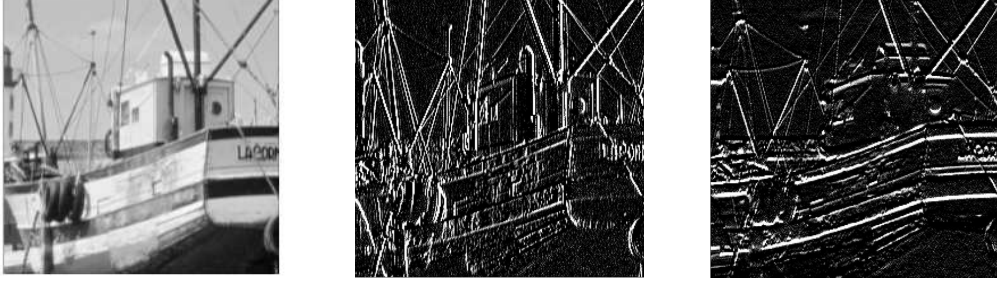


Figure 1: An image, along with its horizontal and vertical discrete directional derivatives.

The discrete gradient transform $\nabla : \mathbb{C}^{N^2} \rightarrow \mathbb{C}^{N \times N \times 2}$ is defined in terms of the directional derivatives,

$$(\nabla \mathbf{x})_{j,k} \stackrel{\text{def}}{=} \begin{cases} ((\mathbf{x}_u)_{j,k}, (\mathbf{x}_v)_{j,k}), & 1 \leq j \leq N-1, \quad 1 \leq k \leq N-1 \\ (0, (\mathbf{x}_v)_{j,k}), & j = N, \quad 1 \leq k \leq N-1 \\ ((\mathbf{x}_u)_{j,k}, 0), & k = N, \quad 1 \leq j \leq N-1 \\ (0, 0), & j = k = N \end{cases}$$

The ℓ_1 -norm of the discrete gradient defines a seminorm for the space \mathbb{C}^{N^2} , often referred to as the *total variation* seminorm and denoted by

$$\|\mathbf{x}\|_{TV} \stackrel{\text{def}}{=} \|\nabla \mathbf{x}\|_1. \quad (6)$$

Due to the gradient sparsity of natural images, it should not be surprising that the total variation minimization program

$$\hat{\mathbf{x}} = \underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{z}\|_{TV} \quad \text{such that} \quad \|\mathcal{M}(\mathbf{z}) - \mathbf{y}\|_2 \leq \varepsilon. \quad (\text{TV})$$

is often used for image reconstruction, in setting of compressed sensing and more broadly in imaging applications such as denoising, deblurring, and inpainting (see e.g. [5, 8, 4, 30, 9, 24, 25, 23, 29, 26, 18, 20, 28] and the references therein). While (TV) is similar to the ℓ_1 -minimization program (L_1), the RIP-based theoretical guarantees for (L_1) do not directly translate to recovery guarantees for (TV) because the gradient map $\mathbf{z} \rightarrow \nabla \mathbf{z}$ is not well-conditioned. In fact, viewed as an invertible operator over mean-zero images, the condition number of the gradient map grows linearly with the signal side-length N . Recovery guarantees for (TV) were nevertheless obtained in [28] for the setting of two-dimensional images $\mathbf{x} \in \mathbb{C}^{N^2}$ by showing that the gradient map is well-conditioned when restricted to signals lying in null space of a matrix with the restricted isometry property:

Theorem A (from [28]). *There are choices of linear operators $\mathcal{M} : \mathbb{C}^{N^2} \rightarrow \mathbb{C}^m$ with $m \approx s \log(N^2/s)$ for which the following holds for any image $\mathbf{x} \in \mathbb{C}^{N^2}$: Given noisy measurements $\mathbf{y} = \mathcal{M}(\mathbf{x}) + \boldsymbol{\xi}$ with noise level $\|\boldsymbol{\xi}\|_2 \leq \varepsilon$, the total-variation minimizing signal*

$$\hat{\mathbf{x}} = \underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{z}\|_{TV} \quad \text{such that} \quad \|\mathcal{M}(\mathbf{z}) - \mathbf{y}\|_2 \leq \varepsilon \quad (7)$$

satisfies the error bound

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq C \log(N^2/s) \left(\frac{\|\nabla \mathbf{x} - (\nabla \mathbf{x})_s\|_1}{\sqrt{s}} + \varepsilon \right), \quad (8)$$

where here and throughout, \mathbf{z}_s denotes the best s -term approximation to the array \mathbf{z} .

In words, the total-variation minimizer estimates \mathbf{x} to within a factor of the noise level and best s -term approximation of its gradient. The bound in (8) is optimal up to the logarithmic factor $\log(N^2/s)$.

The contribution of this paper is to extend Theorem A to multidimensional signals $\mathbf{x} \in \mathbb{C}^{N^d}$ of arbitrary dimension $d \geq 2$. We show that the signal $\hat{\mathbf{x}} \in \mathbb{C}^{N^d}$ of minimal (d -dimensional) total variation seminorm consistent with $m \approx sd \log(N^d)$ appropriately-chosen linear measurements $\mathbf{y} = \mathcal{M}(\mathbf{x}) + \xi$ will approximate \mathbf{x} to within a factor of the noise level and the best s -term approximation to the (d -dimensional) discrete gradient of \mathbf{x} , modulo a single logarithmic factor in the signal dimension N^d . In particular, the results of this paper provide guarantees on the power of total variation minimization in reconstructing three-dimensional digital movies, which represent sequences of gradually-changing images and thus have compressible (three-dimensional) discrete gradient.

Our proof rests on extending the Sobolev inequalities for random subspaces from [28] to higher-dimensional signal structures, using bounds of Cohen, Dahmen, Daubechies, and DeVore in [11] on the compressibility of wavelet representations for functions of bounded variation. Unfortunately these bounds, and hence our results for total variation, do not hold in dimension $d = 1$; guarantees on the fidelity of total variation minimization in the one-dimensional setting remains an interesting open problem.

1.2 Organization

The article is organized as follows. In Section 2 we recall relevant background material on the multidimensional total variation seminorm and multidimensional orthonormal wavelet transform. Section 3 states our main result: total variation minimization provides stable signal recovery for signals of arbitrary dimension $d \geq 2$. The proof of this result will occupy the remainder of the paper; in Section 4 we prove that the signal *gradient* is recovered stably, while in Section 5 we pass from stable gradient recovery to stable signal recovery using the strengthened Sobolev inequalities for random subspaces. The proofs of propositions and theorems used along the way are contained in the appendix.

2 Preliminaries for multidimensional signal analysis

The setting for this article is the space \mathbb{C}^{N^d} of multidimensional arrays of complex numbers, consisting of elements

$$\mathbf{x} = (x_\alpha) \in \mathbb{C}^{N^d}, \quad \alpha \equiv (\alpha_1, \alpha_2, \dots, \alpha_d) \in [N]^d,$$

where $[N]^d = \{1, 2, \dots, N\}^d$. The space \mathbb{C}^{N^d} is a Hilbert space using the standard inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{\alpha \in [N]^d} x_\alpha \cdot \bar{y}_\alpha, \tag{9}$$

and is isometric to the subspace Σ_N of $L_2([0, 1]^d)^1$ consisting of functions which are constant over cubes of side length N^{-1} ,

$$\Sigma_N = \left\{ f \in L_2([0, 1]^d), f(u) = f_\alpha, \quad \frac{\alpha_i - 1}{N} \leq u_i < \frac{\alpha_i}{N} \right\}. \quad (10)$$

For $\mathbf{x} = (x_\alpha) \in \mathbb{C}^{N^d}$, the isometry is provided by identifying $f_\alpha = f_\alpha^\mathbf{x} = N^{d/2} x_\alpha$. More generally, we denote by $\|\mathbf{x}\|_p = \left(\sum_{\alpha \in [N]^d} |x_\alpha|^p \right)^{1/p}$ the entrywise ℓ_p -norm of the signal \mathbf{x} .

For $\ell = 1, 2, \dots, d$, the discrete derivative of \mathbf{x} in the direction of r_ℓ is the array $\mathbf{x}_{r_\ell} \in \mathbb{C}^{N^{\ell-1} \times (N-1) \times N^{d-\ell}}$ defined component-wise by

$$(x_{r_\ell})_\alpha \stackrel{\text{def}}{=} x_{(\alpha_1, \alpha_2, \dots, \alpha_\ell+1, \dots, \alpha_d)} - x_{(\alpha_1, \alpha_2, \dots, \alpha_\ell, \dots, \alpha_d)}, \quad (11)$$

and we define the d -dimensional discrete gradient transform $\nabla : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^{d \times N^d}$ through its components

$$(\nabla \mathbf{x})_{\ell, \alpha} \stackrel{\text{def}}{=} \begin{cases} (x_{r_\ell})_\alpha, & \alpha_\ell \leq N-1, \\ 0, & \text{else} \end{cases} \quad (12)$$

The d -dimensional total-variation seminorm is the ℓ_1 -norm of the d -dimensional discrete gradient,

$$\|\mathbf{x}\|_{TV} \stackrel{\text{def}}{=} \|\nabla \mathbf{x}\|_1. \quad (13)$$

A linear operator $\mathcal{A} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^p$ can be represented as a sequence of multidimensional arrays $\mathcal{A} = (\mathbf{a}_k)$. The linear operation $\mathbf{y} = \mathcal{A}(\mathbf{x})$ has component-wise action

$$y_k = [\mathcal{A}(\mathbf{x})]_k = \langle \mathbf{a}_k, \mathbf{x} \rangle \quad (14)$$

where the inner product between multidimensional arrays is defined in (9). A linear operator $\mathcal{A} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^{N^d}$ can be expressed similarly through its components $y_\alpha = [\mathcal{A}(\mathbf{x})]_\alpha = \langle \mathbf{a}_\alpha, \mathbf{x} \rangle$.

Finally, we introduce the row direct sum operation for concatenating linear operators: if $\mathcal{A} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^{r_1}$ and $\mathcal{B} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^{r_2}$ then $\mathcal{M} = \mathcal{A} \oplus_r \mathcal{B}$ is the linear operator from \mathbb{C}^{N^d} to $\mathbb{C}^{r_1+r_2}$ with component arrays $\mathcal{M} = (\mathbf{m}_k)_{k=1}^{r_1+r_2}$ given by

$$\mathbf{m}_k = \begin{cases} \mathbf{a}_k, & 1 \leq k \leq r_1, \\ \mathbf{b}_{k-r_1}, & 1+r_1 \leq k \leq r_2 \end{cases}$$

Alternatively, the column direct sum operation for concatenating linear operators $\mathcal{A} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^p$ and $\mathcal{B} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^p$ is the linear operator $\mathcal{M} = \mathcal{A} \oplus_c \mathcal{B}$ from $\mathbb{C}^{2 \times N^d}$ to \mathbb{C}^p with component arrays $\mathcal{M} = (\mathbf{m}_k)_{k=1}^p$ given by

$$(\mathbf{m}_k)_{\ell, \alpha} = \begin{cases} (\mathbf{a}_k)_\alpha, & \ell = 1, \\ (\mathbf{b}_k)_\alpha, & \ell = 2 \end{cases}$$

¹Recall that $f \in L_2(Q)$ if $\int_Q |f(u)|^2 du < \infty$, and $L_2(Q)$ is a Hilbert space equipped with the inner product $\langle f, g \rangle = \int_Q f(u) \cdot \bar{g}(u) du$

2.1 The multidimensional Haar wavelet transform

The Haar wavelet transform provides a sparsifying basis for natural signals such as images and movies, and is closely related to the discrete gradient. The Haar transform plays an important role in our analysis for passing from the ℓ_1 -theory of compressed sensing to theory for total variation minimization. For a comprehensive introduction to wavelets, we refer the reader to [15].

The (continuous) multidimensional Haar wavelet basis is derived from a tensor-product representation of the univariate Haar basis, which forms an orthonormal system for square-integrable functions on the unit interval and consists of the constant function

$$h^0(u) = \begin{cases} 1 & 0 \leq t < 1, \\ 0, & \text{otherwise,} \end{cases}$$

the step function

$$h^1(u) = \begin{cases} 1 & 0 \leq t < 1/2, \\ -1 & 1/2 \leq t < 1, \end{cases}$$

and dyadic dilations and translations of the step function,

$$h_{j,k}(t) = 2^{j/2} h^1(2^j t - k); \quad j \in \mathbb{N}, \quad 0 \leq k < 2^j. \quad (15)$$

The Haar basis for the higher dimensional space $L_2(Q)$ of square-integrable functions on the unit cube $Q = [0, 1]^d$ consists of tensor-products of the univariate Haar wavelets. Concretely, for $V = \{0, 1\}^d - \{0\}^d$ and $e = (e_1, e_2, \dots, e_d) \in V$, we define the multivariate functions

$$h^e(u) = \prod_{e_i} h^{e_i}(u_i).$$

The orthonormal Haar system on $L_2(Q)$ is then comprised of the constant function along with all functions of the form

$$h_{j,k}^e(u) = 2^{jd/2} h^e(2^j u - k), \quad e \in V, \quad j \geq 1, \quad k \in \mathbb{Z}^d \cap 2^j Q. \quad (16)$$

The *discrete* multidimensional Haar transform is derived from the continuous construction via the isometric identification (10) between \mathbb{C}^{N^d} and $\Sigma_N \subset L_2(Q)$: defining

$$\mathbf{h}_{j,k,e}(\alpha) = N^{-d/2} h_{j,k}^e(\alpha/N), \quad \alpha \in [N]^d, \quad (17)$$

the matrix product computing the discrete Haar transform can be expressed as $\mathcal{H}\mathbf{x} = (\langle \mathbf{h}_{j,k,e}, \mathbf{x} \rangle)_{j,k,e}$. Note that with this normalization, the transform is orthonormal.

2.2 Gradient versus wavelet sparsity

Since the multivariate Haar transform $\mathcal{H} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^{N^d}$ is orthonormal, standard results in compressed sensing (e.g. [8]) guarantee that by minimizing the ℓ_1 -norm of the conjugate transpose or inverse Haar transform,

$$\tilde{\mathbf{x}} = \underset{\mathbf{z}}{\operatorname{argmin}} \|\mathcal{H}^*(\mathbf{z})\|_1 \quad \text{such that} \quad \|\mathcal{M}(\mathbf{z}) - \mathbf{y}\|_2 \leq \varepsilon \quad (\ell_1\text{-Wav})$$

a multidimensional signal can be reconstructed from $m \geq Cs \log(N^d/s)$ measurements to within a factor of its best s -term approximation in the Haar basis. A straightforward calculation verifies that

a signal which is s -sparse with respect to the discrete gradient is $s \log(N)$ -sparse with respect to the Haar transform. Moreover, this relationship between gradient and wavelet sparsity is stable, as can be seen from the following corollary of results from [11] on the decay of wavelet representations associated to functions of bounded variation:

Proposition 2 (Corollary of Theorem 1.1 from [11]). *There is a universal constant $C > 0$ such that the following holds for any $\mathbf{x} \in \mathbb{C}^{N^d}$ in dimension $d \geq 2$: if the Haar transform coefficients $\mathbf{c} = \mathcal{H}(\mathbf{x})$ are partitioned by their support into blocks $\mathbf{c}_{j,k} = (\langle \mathbf{h}_{j,k,e}, \mathbf{x} \rangle)_{e \in V}$ of cardinality $|\mathbf{c}_{j,k}| = 2^d - 1$, then the coefficient block of k th largest ℓ_2 -norm, denoted by $\mathbf{c}_{(k)}$, has ℓ_2 -norm bounded by*

$$\|\mathbf{c}_{(k)}\|_2 \leq C \frac{\|\mathbf{x}\|_{TV}}{k \cdot 2^{d/2-1}}.$$

Proposition 2, whose derivation from Theorem 1.1 of [11] is outlined in the appendix, suggests that $(\ell_1\text{-Wav})$ should provide reasonably accurate reconstructions for natural signals in the context of compressed sensing.

However, as illustrated in Figure 2 below, the minimizing solutions to $(\ell_1\text{-Wav})$ are known to produce high-frequency artifacts; empirically, the minimizers by total variation tend to produce better reconstructions.

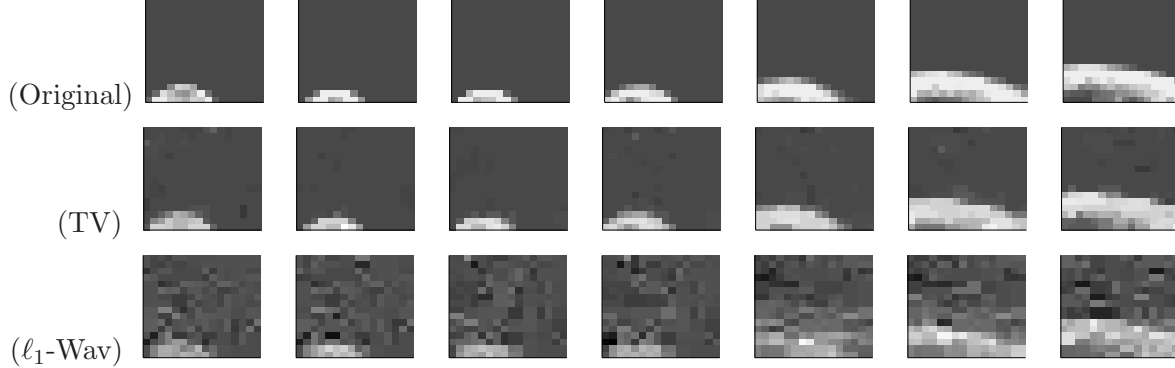


Figure 2: Frame-by-frame comparison between (TV) and $(\ell_1\text{-WAV})$ reconstruction on a $16 \times 16 \times 16$ -pixel MRI movie from $\mathbf{m} = \mathbf{512}$ i.i.d. Gaussian measurements.

Theoretical justification for the good empirical performance of total variation minimization is clearly desirable, and this will occupy the remainder of the present article.

3 The main result

Our main result concerns near-optimal recovery guarantees for multidimensional total variation minimization from compressed measurements. Recall that a linear operator $\mathcal{A} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^p$ is said to have the restricted isometry property (RIP) of order s and level $\delta \in (0, 1)$ when

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|\mathcal{A}(\mathbf{x})\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2 \quad \text{for all } s\text{-sparse } \mathbf{x} \in \mathbb{C}^{N^d}. \quad (18)$$

A linear operator $\mathcal{A} = (\mathbf{a}_k) : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^p$ satisfies the RIP if and only if the $p \times N^d$ matrix A whose k th row consists of the unraveled entries of the k th multidimensional array \mathbf{a}_k satisfies the classical RIP, (1), and so without loss we treat both definitions of the RIP as equivalent.

Analogous to the standard theory of compressed sensing, a lower bound on the optimal recovery rate for an estimate $\hat{\mathbf{x}} = \Delta(\mathcal{M}(\mathbf{x}))$ as a function of $m \gtrsim s \log(N^d/s)$ nonadaptive linear measurements of \mathbf{x} is given by

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \lesssim \frac{1}{d} \left(\frac{\|\nabla \mathbf{x} - (\nabla \mathbf{x})_s\|_1}{\sqrt{s}} + \varepsilon \right), \quad (19)$$

where $(\nabla \mathbf{x})_s$ is the best s -sparse approximation to the discrete gradient $\nabla \mathbf{x}$. Indeed, if a better reconstruction rate were possible, then because $\|\nabla(\hat{\mathbf{x}} - \mathbf{x})\|_2 \leq 2d\|\hat{\mathbf{x}} - \mathbf{x}\|_2$, we would arrive at a better reconstruction rate for $\|\nabla \hat{\mathbf{x}} - \nabla \mathbf{x}\|_2$ than that given by the optimal rate (3). Here we have used the notation $u \gtrsim v$ (analogously $u \lesssim v$) to indicate that there exists some absolute constant $C > 0$ such that $u \geq Cv$ ($u \leq Cv$).

For our main result it will be convenient to define for a multidimensional array $\mathbf{a} \in \mathbb{C}^{N^{\ell-1} \times (N-1) \times N^{d-\ell}}$ the associated arrays $\mathbf{a}_{0_\ell} \in \mathbb{C}^{N^d}$ and $\mathbf{a}^{0_\ell} \in \mathbb{C}^{N^d}$ obtained by concatenating a block of zeros to the beginning and end of \mathbf{a} oriented in the ℓ th direction:

$$(a^{0_\ell})_\alpha = \begin{cases} 0, & \alpha_\ell = 1 \\ a_{\alpha_1, \dots, \alpha_{\ell-1}, \dots, \alpha_d}, & 2 \leq \alpha_\ell \leq N \end{cases} \quad (20)$$

and

$$(a_{0_\ell})_\alpha = \begin{cases} 0, & \alpha_\ell = N \\ a_\alpha, & 1 \leq \alpha_\ell \leq N-1 \end{cases} \quad (21)$$

The following lemma relating gradient measurements with \mathbf{a} to signal measurements with \mathbf{a}^{0_ℓ} and \mathbf{a}_{0_ℓ} can be verified by direct algebraic manipulation and thus the proof is omitted.

Lemma 3. *Given $\mathbf{x} \in \mathbb{C}^{N^d}$ and $\mathbf{a} \in \mathbb{C}^{N^{\ell-1} \times (N-1) \times N^{d-\ell}}$,*

$$\langle \mathbf{a}, \mathbf{x}_{r_\ell} \rangle = \langle \mathbf{a}^{0_\ell}, \mathbf{x} \rangle - \langle \mathbf{a}_{0_\ell}, \mathbf{x} \rangle,$$

where the directional derivative \mathbf{x}_{r_ℓ} is defined in (4).

For a linear operator $\mathcal{A} = (\mathbf{a}_k) : \mathbb{C}^{N^{\ell-1} \times (N-1) \times N^{d-\ell}} \rightarrow \mathbb{C}^m$ we define the operators $\mathcal{A}^{0_\ell} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^m$ and $\mathcal{A}_{0_\ell} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^m$ as the sequences of arrays $(\mathbf{a}_k^{0_\ell})_{k=1}^m$ and $(\mathbf{a}_{0_k})_{k=1}^m$, respectively. From Lemma 22, $\mathcal{A}(\mathbf{x}_{r_\ell}) = \mathcal{A}^{0_\ell}(\mathbf{x}) - \mathcal{A}_{0_\ell}(\mathbf{x})$.

We are now prepared to state our main result which shows that total variation minimization yields stable recovery of N^d -dimensional signals from RIP measurements.

Main Theorem. *Let $N = 2^n$. Fix integers p and q , and let $\mathcal{H} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^{N^d}$ be the orthonormal Haar wavelet transform, and let $\mathcal{A} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^p$ be such that the composite operator $\mathcal{A}\mathcal{H}^{-1} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^p$ has the restricted isometry property of order $2s$ and level $\delta < 1$. Let $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_d$ with $\mathcal{B}_j : \mathbb{C}^{N^{d-1}(N-1)} \rightarrow \mathbb{C}^q$ be such that $\mathcal{B} = \mathcal{B}_1 \oplus_c \mathcal{B}_2 \oplus_c \dots \oplus_c \mathcal{B}_d : \mathbb{C}^{N^{d-1}(N-1)} \rightarrow \mathbb{C}^{dq}$ has the restricted isometry property of order $5s$ and level $\delta < 1/3$. Set $m = 2dq + p$, and consider the linear operator $\mathcal{M} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^m$ given by*

$$\mathcal{M} = \mathcal{A} \oplus_r [\mathcal{B}_1]^{0_1} \oplus_r [\mathcal{B}_1]_{0_1} \oplus_r \dots \oplus_r [\mathcal{B}_\ell]^{0_\ell} \oplus_r [\mathcal{B}_\ell]_{0_\ell} \oplus_r \dots \oplus_r [\mathcal{B}_d]^{0_d} \oplus_r [\mathcal{B}_d]_{0_d}. \quad (22)$$

The following holds for all $\mathbf{x} \in \mathbb{C}^{N^d}$: From noisy measurements $\mathbf{y} = \mathcal{M}(\mathbf{x}) + \boldsymbol{\xi}$ with noise level $\|\boldsymbol{\xi}\|_2 \leq \varepsilon$, the solution to the total variation minimization program,

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_{TV} \quad \text{such that} \quad \|\mathcal{M}(\mathbf{z}) - \mathbf{y}\|_2 \leq \varepsilon \quad (23)$$

satisfies:

- i.* $\|\nabla(\mathbf{x} - \hat{\mathbf{x}})\|_2 \lesssim \frac{\|\nabla\mathbf{x} - (\nabla\mathbf{x})_s\|_1}{\sqrt{s}} + \sqrt{d}\varepsilon,$
- ii.* $\|\mathbf{x} - \hat{\mathbf{x}}\|_{TV} \lesssim \|\nabla\mathbf{x} - (\nabla\mathbf{x})_s\|_1 + \sqrt{sd}\varepsilon,$
- iii.* $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \lesssim \log(N^d) \left(\frac{\|\nabla\mathbf{x} - (\nabla\mathbf{x})_s\|_1}{\sqrt{s}} + \sqrt{d}\varepsilon \right).$

Remarks.

1. A number of $m \gtrsim sd \log(N^d)$ i.i.d. and properly normalized Gaussian measurements can be used to construct the measurement operator \mathcal{M} which, with high probability, satisfies the required RIP conditions of the theorem [1, 34]. From this number m of measurements, the error guarantees *i* and *ii* are optimal up to the factor of \sqrt{d} on the noise level, and the error guarantee *iii* is optimal up to logarithmic factors in the signal dimension N^d .

2. When $d = 2$, the main theorem recovers the total variation stability guarantee of [28] up to a $\log(1/s)$ term.

3. The requirement of sidelength $N = 2^n$ is not an actual restriction, as signals with arbitrary side-length N can be extended via reflections across each dimension to a signal of side-length $N = 2^n$ without increasing the total variation by more than a factor of 2^d . This requirement again seems to be only an artifact of the proof and one need not perform such changes in practice.

We now turn to the proof of the main theorem. We follow the method of proof introduced in [28], proving stable gradient recovery and then translating these guarantees to stable signal recovery via Sobolev inequalities for incoherent subspaces.

4 Stable gradient recovery

In this section we prove statements (*i*) and (*ii*) of the main theorem concerning stable gradient recovery, using standard results in the ℓ_1 theory of compressed sensing combined with a summation by parts “trick” provided by Lemma 3.

Recall that when a signal obeys a tube and cone constraint we can bound the norm of the entire signal, as in [8]. We refer the reader to Section A.1 of [28] for a complete proof.

Proposition 4. *Suppose that \mathcal{A} is a linear operator satisfying the restricted isometry property of order $5s$ and level $\delta < 1/3$, and suppose that the signal \mathbf{v} satisfies a tube constraint*

$$\|\mathcal{A}(\mathbf{v})\|_2 \lesssim \varepsilon.$$

Suppose further that for a subset S of cardinality $|S| = s$ (with complement S^c), \mathbf{v} satisfies a cone-constraint

$$\|\mathbf{v}_{S^c}\|_1 \leq \|\mathbf{v}_S\|_1 + \xi. \tag{24}$$

Then

$$\|\mathbf{v}\|_2 \lesssim \frac{\xi}{\sqrt{s}} + \varepsilon \tag{25}$$

and

$$\|\mathbf{v}\|_1 \lesssim \xi + \sqrt{s}\varepsilon. \tag{26}$$

Using Proposition 4 and the RIP assumptions on the operator \mathcal{B} , stable gradient recovery (i) and (ii) reduce to proving that the discrete gradient of the residual signal error satisfies the tube and cone constraints.

Proof. (Main Theorem, statements (i) and (ii).) Let $\mathbf{v} = \mathbf{x} - \hat{\mathbf{x}}$ be the residual error. Let $\mathcal{B} = (\mathbf{b}_k)$. Then we have

Cone Constraint. Let S denote the support of the best s -sparse approximation to $\nabla \mathbf{x}$ and let S_c be the complement of S . Since $\hat{\mathbf{x}} = \mathbf{x} - \mathbf{v}$ is the minimizer of (TV) and \mathbf{x} satisfies the feasibility constraints in (TV), $\|\nabla \hat{\mathbf{x}}\|_1 \leq \|\nabla \mathbf{x}\|_1$ and by the reverse triangle inequality,

$$\begin{aligned} \|(\nabla \mathbf{x})_S\|_1 - \|(\nabla \mathbf{v})_S\|_1 - \|(\nabla \mathbf{x})_{S^c}\|_1 + \|(\nabla \mathbf{v})_{S^c}\|_1 \\ \leq \|(\nabla \mathbf{x})_S + (\nabla \mathbf{v})_S\|_1 + \|(\nabla \mathbf{x})_{S^c} + (\nabla \mathbf{v})_{S^c}\|_1 \\ = \|\nabla \hat{\mathbf{x}}\|_1 \\ \leq \|\nabla \mathbf{x}\|_1 \\ = \|(\nabla \mathbf{x})_S\|_1 + \|(\nabla \mathbf{x})_{S^c}\|_1 \end{aligned}$$

This yields the cone constraint

$$\|(\nabla \mathbf{v})_{S^c}\|_1 \leq \|(\nabla \mathbf{v})_S\|_1 + 2\|(\nabla \mathbf{x})_{S^c}\|_1.$$

Tube constraint. Recall that $\mathbf{v} = \mathbf{x} - \hat{\mathbf{x}}$. Since both \mathbf{x} and $\hat{\mathbf{x}}$ are feasible solutions to (TV), the triangle inequality gives

$$\begin{aligned} \|\mathcal{M}(\mathbf{v})\|_2^2 &\leq \|\mathcal{M}(\mathbf{x}) - \mathbf{y}\|_2^2 + \|\mathcal{M}(\hat{\mathbf{x}}) - \mathbf{y}\|_2^2 \\ &\leq 4\varepsilon^2 \end{aligned}$$

By Lemma 3, we have for each component operator \mathcal{B}_j ,

$$\mathcal{B}_j(\mathbf{v}_{r_j}) = [\mathcal{B}_j]^{0_j}(\mathbf{v}) - [\mathcal{B}_j]_{0_j}(\mathbf{v}) \quad (27)$$

Then $\mathcal{B}(\nabla \mathbf{v}) = \sum_{j=1}^d \mathcal{B}_j(\mathbf{v}_{r_j})$, (where we assume that $\nabla \mathbf{v}$ is ordered appropriately) and

$$\begin{aligned} \|\mathcal{B}(\nabla \mathbf{v})\|_2^2 &= \left\| \sum_{j=1}^d \mathcal{B}_j(\mathbf{v}_{r_j}) \right\|_2^2 \\ &\leq d \sum_{j=1}^d \|\mathcal{B}_j(\mathbf{v}_{r_j})\|_2^2 \\ &\leq 2d \sum_{j=1}^d \left(\|[\mathcal{B}_j]^{0_j}(\mathbf{v})\|_2^2 + \|[\mathcal{B}_j]_{0_j}(\mathbf{v})\|_2^2 \right) \\ &\leq 2d \|\mathcal{M}(\mathbf{v})\|_2^2 \\ &\leq 2d\varepsilon^2. \end{aligned} \quad (28)$$

In light of Proposition 4 this completes the proof. □

Remark 5. The component operator \mathcal{A} from the main theorem was not used at all in deriving properties (i) and (ii); on the other hand, only the measurements in \mathcal{A} will be used to derive property (iii) from (i) and (ii).

5 A Sobolev inequality for incoherent subspaces

The purpose of this section is to derive the following bound for signals lying near the null space of an incoherent matrix.

Theorem 6 (Sobolev inequality for incoherent subspaces). *Let $d \geq 2$ and let $N = 2^n$. Let $\mathcal{H} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^{N^d}$ be the multivariate Haar wavelet transform. Let $\mathcal{B} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^m$ be a linear map with the property that $\mathcal{B}\mathcal{H}^{-1} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^m$ satisfies the restricted isometry property of order $2s$ and level $\delta < 1$. Then there is a universal constant $C > 0$ such that for any signal $\mathbf{x} \in \mathbb{C}^{N^d}$ satisfying the tube constraint $\|\mathcal{B}(\mathbf{x})\|_2 \leq \varepsilon$,*

$$\|\mathbf{x}\|_2 \leq C \left(\frac{\|\mathbf{x}\|_{TV}}{\sqrt{s}} \right) \log(N^d) + \varepsilon. \quad (29)$$

Note that Theorem 6 admits various corollaries for various families of random matrices with restricted isometries. For subgaussian random matrices, we have using the results in [1]

Corollary 7. *Let $\mathcal{B} : \mathbb{C}^{N^d} \rightarrow \mathbb{C}^m$ be a linear map realizable as an $N^d \times m$ matrix whose entries are mean-zero i.i.d. Gaussian random variables. Then with probability exceeding $1 - e^{-cm}$, the following bound holds for any $\mathbf{x} \in \mathbb{C}^{N^d}$ lying in the null-space of \mathcal{B} :*

$$\|\mathbf{x}\|_2 \lesssim \left(\frac{\|\mathbf{x}\|_{TV}}{\sqrt{m}} \right) [\log(N^d)]^2. \quad (30)$$

The proof of Theorem 6 goes as follows: as revealed through the proof of Proposition 4, any signal in the null space of an RIP matrix must be relatively *flat*, in that the norm of its best s -term approximation is bounded by the norm of the remaining terms. At the same time, Proposition 2 implies that the sequence of wavelet block-coefficient norms $\|\mathbf{c}_{(k)}\|_2$ in the null space of $\mathcal{B}\mathcal{H}^{-1}$ must be sufficiently compressible and bounded by the total-variation of \mathcal{B} . Combining these properties — flatness and tail-compressibility — produces the Sobolev-type inequality (6).

Proof of Theorem 6. Let $\mathbf{c} = \mathcal{H}(\mathbf{v}) \in \mathbb{C}^{N^d}$ represent the Haar transform of the signal error $\mathbf{v} = \mathbf{x} - \hat{\mathbf{x}}$. Suppose without loss that the desired sparsity level s is either smaller than $2^d - 1$ or a positive multiple of $2^d - 1$, and write $s = p(2^d - 1)$ where either $p \in \mathbb{N}$ or $p \in (0, 1)$ (for arbitrary $s \in \mathbb{N}$, we could consider $s' = \lceil s/(2^d - 1) \rceil$ which satisfies $s' \leq 2s$).

Let $S = S_0 \subset [N]^d$ be the set of s largest-magnitude entries of \mathbf{c} , let S_1 be the set of s largest-magnitude entries of \mathbf{c} in $[N]^d \setminus S_0$, and so on. Note that \mathbf{c}_S and similar expressions below can have both the meaning of restricting \mathbf{c} to the indices in S as well as being the array whose entries are set to zero outside S .

By definition, $\|\mathbf{c}_{S_0}\|_1$ is at least as large as $\|\mathbf{c}_\Omega\|_1$ for any other $\Omega \subset [N]^d$ of cardinality s . Consequently, $\|\mathbf{c}_{S_0^c}\|_1$ is *smaller* than $\|\mathbf{c}_{\Omega^c}\|_1$ for any other $\Omega \subset [N]^d$ of cardinality s . Now, recall the alternative decomposition of \mathbf{c} from Proposition 2 into blocks $\mathbf{c}_{(k)}$ of cardinality $2^d - 1$, grouped

according to the support of the corresponding wavelets. Because $s = p(2^d - 1)$ for either $p \in \mathbb{N}$ or $p \in (0, 1)$,

$$\begin{aligned}
\|\mathbf{c}_{S_0^c}\|_1 &= \sum_{k \geq 1} \|\mathbf{c}_{S_k}\|_1 \\
&\leq \sum_{j \geq p+1} \|\mathbf{c}_{(j)}\|_1 \\
&\leq (2^d - 1)^{1/2} \sum_{j \geq p+1} \|\mathbf{c}_{(j)}\|_2 \\
&\lesssim \|\mathbf{v}\|_{TV} \sum_{\ell=p+1}^{N^d} \frac{1}{\ell} \quad \text{by Proposition 2} \\
&\lesssim \|\mathbf{v}\|_{TV} \log(N^d),
\end{aligned} \tag{31}$$

where the last inequality follows from properties of the geometric summation.

We use a similar procedure to bound the ℓ_2 -norm of the residual,

$$\begin{aligned}
\|\mathbf{c}_{S_0^c}\|_2^2 &\lesssim \sum_{j \geq p+1} \|\mathbf{c}_{(j)}\|_2^2 \\
&\lesssim \frac{\|\mathbf{v}\|_{TV}^2}{2^d} \sum_{\ell=p+1}^{N^d} \frac{1}{\ell^2} \\
&\lesssim \frac{(\|\mathbf{v}\|_{TV})^2}{2^d \max(1, p)} \\
&\lesssim \frac{(\|\mathbf{v}\|_{TV})^2}{s}.
\end{aligned} \tag{32}$$

Then, $\|\mathbf{c}_{S_0^c}\|_2 \lesssim \|\mathbf{v}\|_{TV} / \sqrt{s}$.

By assumption, \mathbf{v} satisfies the tube constraint $\|\mathcal{B}(\mathbf{v})\|_2 \leq \varepsilon$ and $\mathcal{B}\mathcal{H}^{-1}$ satisfies the restricted isometry property. We conclude that

$$\begin{aligned}
\varepsilon &\geq \|\mathcal{B}(\mathbf{v})\|_2 = \|\mathcal{B}\mathcal{H}^{-1}(\mathbf{c})\|_2 \\
&\geq \|\mathcal{B}\mathcal{H}^{-1}(\mathbf{c}_{S_0} + \mathbf{c}_{S_1})\|_2 - \sum_{k=2}^r \|\mathcal{B}\mathcal{H}^{-1}(\mathbf{c}_{S_k})\|_2 \\
&\geq (1 - \delta) \|\mathbf{c}_{S_0} + \mathbf{c}_{S_1}\|_2 - (1 + \delta) \sum_{k=2}^r \|\mathbf{c}_{S_k}\|_2 \\
&\geq (1 - \delta) \|\mathbf{c}_{S_0}\|_2 - (1 + \delta) \frac{1}{\sqrt{s}} \|\mathbf{c}_{S_0^c}\|_1,
\end{aligned} \tag{33}$$

the last inequality holding because the magnitude of each entry in the array \mathbf{c}_{S_k} is smaller than the average magnitude of the entries in the array $\mathbf{c}_{S_{k-1}}$. Along with the tail bound (31), we can then conclude that, up to a constant in the restricted isometry level δ ,

$$\|\mathbf{c}_{S_0}\|_2 \lesssim \varepsilon + \log(N^d) \left(\frac{\|\mathbf{v}\|_{TV}}{\sqrt{s}} \right). \tag{34}$$

Combining this bound with the ℓ_2 -tail bound (32) and recalling that the Haar transform $\mathcal{H} : N^d \rightarrow N^d$ is an isometry,

$$\|\mathbf{v}\|_2 = \|\mathcal{H}^{-1}\mathbf{c}\|_2 = \|\mathbf{c}\|_2 \leq \|\mathbf{c}_{S_0}\|_2 + \|\mathbf{c}_{S_0^c}\|_2 \lesssim \varepsilon + \log(N^d) \left(\frac{\|\mathbf{v}\|_{TV}}{\sqrt{s}} \right), \quad (35)$$

which completes the proof. \square

5.1 Proof of the Main Theorem

Because we proved the bounds (i) and (ii) from the main theorem concerning stable gradient recovery in Section 4, it remains only to prove the signal recovery error bound (iii).

By feasibility of both \mathbf{x} and $\hat{\mathbf{x}}$ for the constraints in the total variation minimization program, the signal error $\mathbf{v} = \mathbf{x} - \hat{\mathbf{x}}$ obeys the tube-constraint $\|\mathcal{A}(\mathbf{v})\|_2 \leq 2\varepsilon$. Applying Theorem 6 and the total variation bound (ii) yields

$$\begin{aligned} \|\mathbf{x} - \hat{\mathbf{x}}\|_2 &= \|\mathbf{v}\|_2 \\ &\lesssim \varepsilon + \log(N^d) \left(\frac{\|\mathbf{v}\|_{TV}}{\sqrt{s}} \right) \\ &\lesssim \varepsilon + \log(N^d) \left(\frac{\|\nabla \mathbf{x} - (\nabla \mathbf{x})_s\|_1 + \sqrt{s}\sqrt{d}\varepsilon}{\sqrt{sd}} \right) \\ &\lesssim \log(N^d) \left(\frac{\|\nabla \mathbf{x} - (\nabla \mathbf{x})_s\|_1}{\sqrt{s}} + \sqrt{d}\varepsilon \right). \end{aligned}$$

The proof completes.

A Derivation of Proposition 2

Recall that the space $L_p(\Omega)$ ($1 \leq p < \infty$) for $\Omega \subset \mathbb{R}^d$ consists of all functions f satisfying

$$\|f\|_{L_p(\Omega)} = \left(\int_{\Omega} |f(u)|^p du \right)^{1/p} < \infty.$$

The space $BV(\Omega)$ of functions of bounded variation over the unit cube $Q = [0, 1]^d$ is often used as a continuous model for natural images. Recall that a function $f \in L_1(Q)$ has finite bounded variation if and only if its distributional gradient is a finite measure, and this measure generates the BV seminorm $|f|_{BV(Q)}$. More precisely,

Definition 8. For a vector $\mathbf{v} \in \mathbb{R}^d$, we define the difference operator $\Delta_{\mathbf{v}}$ in the direction of \mathbf{v} by

$$\Delta_{\mathbf{v}}(f, \mathbf{x}) := f(\mathbf{x} + \mathbf{v}) - f(\mathbf{x}).$$

We say that a function $f \in L_1(Q)$ is in $BV(Q)$ if and only if

$$V_Q(f) \stackrel{\text{def}}{=} \sup_{h>0} h^{-1} \sum_{j=1}^d \|\Delta_{h\mathbf{e}_j}(f, \cdot)\|_{L_1(Q(h\mathbf{e}_j))} = \lim_{h \rightarrow 0} h^{-1} \sum_{j=1}^d \|\Delta_{h\mathbf{e}_j}(f, \cdot)\|_{L_1(Q(h\mathbf{e}_j))} < \infty$$

where \mathbf{e}_j denotes the j th coordinate vector. The function $V_Q(f)$ provides a semi-norm for $BV(Q)$:

$$|f|_{BV(Q)} \stackrel{\text{def}}{=} V_Q(f).$$

In particular, piecewise constant functions are in the space $BV(Q)$. For $N = 2^n$ a power of two, recall the following relationship between the total-variation of a multidimensional signal $\mathbf{x} \in \mathbb{C}^{N^d}$ and the bounded variation of its isometric piecewise-constant representation $f \in \Sigma_N \subset L_2(Q)$ as defined in (10):

Lemma 9. *Let $N = 2^n$ for some integral n . Let $\mathbf{x} \in \mathbb{C}^{N^d}$ and let $f \in \Sigma_N$ be its isometric embedding as in (10). Then $|f|_{BV} \leq N^{-d/2+1} \|\mathbf{x}\|_{TV}$.*

Proof. For $h < \frac{1}{N}$,

$$\Delta_{he_k}(f, (\mathbf{u})) = \begin{cases} N^{d/2}(\mathbf{x}_{\ell^{(k)}} - \mathbf{x}_\ell) & \frac{\ell_i}{N} - h \leq u_i \leq \frac{\ell_i}{N}, \\ 0, & \text{else,} \end{cases}$$

where

$$\ell_i^{(k)} = \begin{cases} \ell_i & i \neq k, \\ \ell_i + 1, & i = k. \end{cases}$$

Thus

$$\begin{aligned} |f|_{BV} &= \lim_{h \rightarrow 0} \frac{1}{h} \sum_{k=1}^d \left[\int_0^1 \int_0^1 \dots \int_0^1 |f(\mathbf{u} + he_k) - f(\mathbf{u})| d\mathbf{u} \right] \\ &= \sum_{k=1}^d N^{d/2} \left[\sum_{\ell} \frac{1}{N} |\mathbf{x}_{\ell^{(k)}} - \mathbf{x}_\ell| \right] \\ &\leq N^{-d/2+1} \|\mathbf{x}\|_{TV}. \end{aligned}$$

□

Cohen, Dahmen, Daubechies, and DeVore showed in [11] that the properly normalized sequence of rearranged wavelet coefficients associated to a function $f \in L_2(\Omega)$ of bounded variation is in $weak\text{-}\ell_1$, and its weak- ℓ_1 seminorm is bounded by the function BV seminorm. Using different normalizations to those used in [11] — we use the L_2 -normalization for the Haar wavelets as opposed to the L_1 -normalization — we consider the Haar wavelet coefficients $f_I^e = \langle f, h_I^e \rangle$ and consider the wavelet coefficient block $f_I = (f_I^e)_{e \in E} \in \mathbb{C}^{2^{d-1}}$ associated to those Haar wavelets supported on the dyadic cube I . With this notation, Theorem 1.1 of [11] applied to the Haar wavelet system over $L_2(Q)$ reads

Proposition 10. *Let $d \geq 2$. Then there exists a constant $C > 0$ such that the following holds for all $f \in BV(Q)$. Let the wavelet coefficient block with k th largest ℓ_2 -norm be denoted by $f^{(k)}$, and suppose that this block is associated to the dyadic cube $I_{j,k}$ with side-length 2^{-j} . Then*

$$\|f^{(k)}\|_2 \leq C \frac{2^{j(d-2)/2} |f|_{BV}}{k}.$$

Proposition 2 results by translating Proposition 10 to the discrete setting of \mathbb{C}^{N^d} using the isometry (10) and Lemma 9. We note that a stronger version of this result was provided for the 2-dimensional Haar wavelet basis in [12] and used in the proof of stable image recovery from total-variation minimization in [28].

Acknowledgment

We would like to thank John Doyle, Christina Frederick, and Mark Tygert for helpful improvements and insights.

References

- [1] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin. A simple proof of the Restricted Isometry Property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.
- [2] N. Burq, S. Dyatlov, R. Ward, and M. Zworski. Weighted eigenfunction estimates with applications to compressed sensing. *arXiv:1111.2383*, 2012.
- [3] E. Candès, Y. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Harmon. A.*, 31(1):59–73, 2011.
- [4] E. Candès and J. Romberg. Signal recovery from random projections. In *Proc. SPIE Conference on Computational Imaging III*, volume 5674, pages 76–86. SPIE, 2005.
- [5] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.
- [6] E. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51:4203–4215, 2005.
- [7] E. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.
- [8] E. Candès, T. Tao, and J. Romberg. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
- [9] T.F. Chan, J. Shen, and H.M. Zhou. Total variation wavelet inpainting. *J. Math. Imaging Vis.*, 25(1):107–125, 2006.
- [10] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by Basis Pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1999.
- [11] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore. Harmonic analysis of the space BV. *Revista de Matematica Iberoamericana*, 19(1):235–263, 2003.
- [12] A. Cohen, R. DeVore, P. Petrushev, and H. Xu. Nonlinear approximation and the space $BV(\mathbb{R}^2)$. *Am. J. of Math*, 121:587–628, 1999.
- [13] Compressed sensing webpage. <http://www.dsp.ece.rice.edu/cs/>.
- [14] G. Dantzig and M. Thapa. *Linear Programming*. Springer, New York, NY, 1997.
- [15] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.

- [16] D. Donoho and P. Stark. Uncertainty principles and signal recovery. *SIAM J. on Appl. Math.*, pages 906–931, 1989.
- [17] A. Garnaev and E. Gluskin. On widths of the Euclidean ball. *Sov. Math. Dokl.*, 30:200–204, 1984.
- [18] B. Kai Tobias, U. Martin, and F. Jens. Suppression of MRI truncation artifacts using total variation constrained data extrapolation. *Int. J. Biomedical Imaging*, 2008.
- [19] B. Kashin. The widths of certain finite dimensional sets and classes of smooth functions. *Izvestia*, 41:334–351, 1977.
- [20] S.L. Keeling. Total variation based convex filters for medical imaging. *Appl. Math. Comput.*, 139(1):101–119, 2003.
- [21] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *arxiv:1207.0235*, 2012.
- [22] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.
- [23] Y. Liu and Q. Wan. Total variation minimization based compressive wideband spectrum sensing for cognitive radios. Submitted, 2011.
- [24] M. Lustig, D. Donoho, and J.M. Pauly. Sparse MRI: The application of compressed sensing for rapid MRI imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.
- [25] M. Lustig, D. Donoho, J. Santos, and J. Pauly. Compressed sensing MRI. *IEEE Sig. Proc. Mag.*, 25(2):72–82, 2008.
- [26] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty. An efficient algorithm for compressed MR imaging using total variation and wavelets. In *IEEE Conf. Comp. Vision Pattern Recog.*, 2008.
- [27] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constr. Approx.*, 28(3):277–289, 2008.
- [28] D. Needell and R. Ward. Stable image reconstruction using total variation minimization. Submitted, 2012.
- [29] B. Nett, J. Tang, S. Leng, and G.H. Chen. Tomosynthesis via total variation minimization reconstruction and prior image constrained compressed sensing (PICCS) on a C-arm system. In *Proc. Soc. Photo-Optical Instr. Eng.*, volume 6913. NIH Public Access, 2008.
- [30] S. Osher, A. Solé, and L. Vese. Image decomposition and restoration using total variation minimization and the H^{-1} norm. *Multiscale Model. Sim.*, 1:349–370, 2003.
- [31] H. Rauhut. Compressive Sensing and Structured Random Matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.
- [32] H. Rauhut and R. Ward. Sparse recovery for spherical harmonic expansions. *Proc. SampTA*, 2011.

- [33] H. Rauhut and R. Ward. Sparse Legendre expansions via ℓ_1 -minimization. *Journal of Approximation Theory*, 164:517–533, 2012.
- [34] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.